Calibration Model Maintenance in Melamine Resin Production: Integrating Drift Detection, Smart Sample Selection and Model Adaptation

Ramin Nikzad-Langerodi[a], Edwin Lughofer[a], Carlos Cernuda[b], Thomas Reischer[c], Wolfgang Kantner[c], Marcin Pawliczek[d], Markus Brandstetter[d]

[a]*Department of Knowledge-Based Mathematical Systems, Johannes Kepler University Linz, Austria*
[b]*BCAM - Basque Center for Applied Mathematics, Bilbao, Spain*
[c]*Metadynea, Krems, Austria*
[d]*RECENDT Research Center for Non-Destructive Testing GmbH, Linz, Austria*

Abstract

The physico-chemical properties of Melamine Formaldehyde (MF) based thermosets are largely influenced by the degree of polymerization (DP) in the underlying resin. On-line supervision of the turbidity point by means of vibrational spectroscopy has recently emerged as a promising technique to monitor the DP of MF resins. However, spectroscopic determination of the DP relies on chemometric models, which are usually sensitive to *drifts* caused by instrumental and/or sample-associated changes occurring over time. In order to detect the time point when *drifts* start causing prediction bias, we here explore a universal drift detector based on a faded version of the Page-Hinkley (PH) statistic, which we test in three data streams from an industrial MF resin production process. We employ *committee disagreement* (CD), computed as the variance of model predictions from an ensemble of partial least squares (PLS) models, as a measure for sample-wise prediction uncertainty and use the PH statistic to detect changes in this quantity. We further explore supervised and unsupervised strategies for (semi-)automatic model adaptation upon detection of a drift. For the former, manual reference measurements are requested whenever statistical thresholds on Hotelling's T^2 and/or Q-Residuals are violated. Models are subsequently re-calibrated using weighted partial least squares in order to increase the influence of newer samples, which increases the flexibility when adapting to new (drifted) states. Unsupervised model adaptation is carried out exploiting the dual antecedent-consequent structure of a recently developed fuzzy systems variant

Preprint submitted to Analytica Chimica Acta *September 14, 2018*

Email addresses: ramin.nikzad-langerodi@jku.at (Ramin Nikzad-Langerodi), edwin.lughofer@jku.at (Corresponding Author) (Edwin Lughofer), ccernuda@bcamath.org (Carlos Cernuda), thomas.reischer@metadynea.com (Thomas Reischer), wolfgang.kantner@metadynea.com (Wolfgang Kantner), marcin.pawliczek@recendt.at (Marcin Pawliczek), markus.brandstetter@recendt.at (Markus Brandstetter)

of PLS termed FLEXFIS-PLS. In particular, antecedent parts are updated while maintaining the internal structure of the local linear predictors (i.e. the consequents). We found improved drift detection capability of the CD compared to Hotelling's T^2 and Q-Residuals when used in combination with the proposed PH test. Furthermore, we found that active selection of samples by active learning (AL) used for subsequent model adaptation is advantageous compared to passive (random) selection in case that a drift leads to persistent prediction bias allowing more rapid adaptation at lower reference measurement rates. Fully unsupervised adaptation using FLEXFIS-PLS could improve predictive accuracy significantly for light drifts but was not able to fully compensate for prediction bias in case of significant lack of fit w.r.t. the latent variable space.

Keywords: ensembled PLS, drift detection, committee disagreement, active learning, calibration model maintenance by adaptation, weighted learning, melamine formaldehyde resin

1. Introduction

1.1. Motivation and State-of-the-Art

Melamine formaldehyde (MF) is an important class of thermosetting polymers, which have found wide industrial application as coating and adhesive materials owing to their high thermal- and chemical stability and physical resistance [1]. The properties of MF are largely determined by the curing behavior and the degree of polymerization and cross-linking of the final product [2]. It is therefore of pivotal interest to supervise the degree of polymerization (DP) during MF resin production and to determine the optimal time point when the polymerization reaction should be stopped. The water tolerance method and determination of free melamine concentration by liquid chromatography (LC) are well established and widely used to assess the DP of MF resins but produce significant cost in terms of material and human resources [3][4]. In addition, the delay between sample collection and retrieval of the critical process information harbors the risk of missing the optimal time point when the reaction has reached the desired DP [5]. In practice, MF condensation processes are therefore often run below the theoretically possible speed in order to minimize this risk at the expense of suboptimal allocation of resources.

Non-destructive estimation of free melamine content in MF resins by vibrational spectroscopy has been demonstrated previously and holds promise for high-frequency in-line monitoring of MF condensation [3]. Pawlicek *et al.* have recently established an online monitoring and control system for MF resins based on Fourier-transform near infrared (FT-NIR) spectroscopy and demonstrated the feasibility of accurate spectroscopic estimation of the turbidity point during the condensation process [4]. However, it is widely appreciated that chemometric models, usually required for spectroscopic estimation of critical process parameters, often have a limited time-span during which they deliver reliable predictions on new samples (e.g. spectra) and thus need regular supervision and maintenance [6]. The breakdown of model reliability might occur either due to instrumental changes, changes in environmental conditions during spectra acquisition or changes in sample composition (e.g. due to changing raw material composition) [7]. In any case, the relation between FT-NIR spectra and the target values eventually gets altered by such *drifts*, which might ultimately lead to prediction bias.

Maintenance of chemometric models can be roughly divided in two tasks: i) Detection of the time point when any drift starts causing prediction bias (i.e. drift detection) and ii) model adaptation once a drift has been detected. Timely detection of drifts is crucial to avoid cost-intensive manufacturing errors, defective production or even complete shutdown of the process [8]. Re-calibration of chemometric models according to a static schedule using regularly collected reference measurements could eventually compensate for the occurrence of drifts to some extent, but requires time-intensive modeling/validation cycles and disproportional efforts in terms of reference analytics. We have therefore recently proposed a self-adaptive calibration modelling strategy employing non-linear fuzzy PLS models in order to automate off-line re-calibration and validation [5]. However, despite implicit drift handling [9], full supervised model adaptation employed in [5] and [10] [11] require a high number of reference measurements and is thus of little practical relevance. In an attempt to reduce the amount of reference measurements Cernuda *et al.* [12] proposed an active learning (AL) based approach to request (manual) reference measurements for new samples when the corresponding spectra violate upper control limits for Hotelling's T^2

statistic or Q-residuals. Therein, the authors employed a sliding window approach featuring user-defined thresholding of the rate at which reference measurement should be conducted. However, despite dramatic reduction of reference measurements, the method performs permanent sample selection and model re-calibration which may lead to superfluous reference measurements. In addition, employing a fixed window size might be problematic as large windows typically lack the ability to keep pace with the dynamics induced by the drift while important sources of process variability might be disregarded when using small ones. Other works on self-adaptive (evolving) chemometric models have been proposed in [10] [11], but also require the full knowledge of target measurements and are thus not applicable in our context. Active learning for off-line chemometric calibration has been proposed in [13] [14]. However, these approaches are not directly applicable for online model adaptation. For this reason, we here combine active learning along with explicit drift detection: Initially trained chemometric models are updated according to a complete on-line, single-pass active learning concept only once a drift has been detected. In addition, we also explore unsupervised model adaptation exploiting the dual antecedent-consequent structure of the recently proposed fuzzy systems variant of PLS (FLEXFIS-PLS) in order to reduce prediction bias.

1.2. Summary of Our Approach

We use linear and non-linear versions of (ensembled) partial least squares (PLS) in order to establish initial calibration models for the determination of the turbidity point of MF resin during polymerization. For the latter we use a previously described fuzzy-inference based PLS variant termed FLEXFIS-PLS [15] to account for eventual non-linearities with the additional possibility to update initially trained models in an unsupervised fashion. A novel type of drift detector based on a faded version of the Page-Hinkley (PH) statistic [16] is introduced in order to determine the time point when predictions start becoming unreliable and the model adaptation engine should be turned on. In particular, we compute sample-wise prediction uncertainty as the variance of predictions from an ensemble of PLS models, i.e. the *committee disagreement (CD)*, and incrementally re-estimate the distribution of the PH statistic on the CD at every time instance [17]. This yields adaptive control limits that are used to decide if a change has occured. Model updates are triggered

only once these adaptive control limits are violated for the first time, which is an essential difference to the previous work in [12], where calibration models are *permanently* updated over sliding windows.

We investigate different strategies for incremental updates of initially trained chemometric models: i) Active learning in single-pass on-line mode based on violation of the well-known Hotelling's T^2- and Q-residual limits followed by automatic model re-calibration involving adaptive training set re-weighting (giving newer samples higher emphasis) and ii) unsupervised adaptation of non-linear FLEXFIS-PLS models involving incremental adaptation of antecedents, which affects only the position and spreads of the underlying fuzzy rules while maintaining the local linear input-output relationships in the consequents. The rest of the paper is organized as follows: Section 2 provides a brief description of the melamine resin production process and spectroscopic data acquisition. Section 3 develops the theory behind PLS (3.2.1) and its non-linear variant FLEXFIS-PLS (3.2.2) followed by a description of how these models are used for (un-) supervised model adaptation in section 3.3. Active sample selection in single-pass mode is introduced in section 3.3.2. A novel drift detector combining committee disagreement and a modified variant of the Page-Hinkley test is presented in section 3.4. Section 4 describes the experimental setup including subsections describing data pre-treatment, model parametrization and method evaluation. A comparison of the proposed drift detector and different state-of-the-art (SoA) model diagnostic tools (sample-specific standard errors, Hotelling's T^2 and Q-residuals) is presented in section 5 followed by an evaluation of the proposed strategies for (automated) model adaptation (Sections 5.2 and 5.3). Conclusion are drawn in Section 6.

2. Description of the Process and Data Acquisition

The applied FT-NIR process spectrometer is a commercially available device (i-RED Infrarot Systeme, Austria). It is connected with three immersion probes through fiber optic equipment and has been previously implemented at the MF resin production plant allowing parallel monitoring of three reactors [18]. The optics of the spectrometer contain a monolithic Michelson interferometer

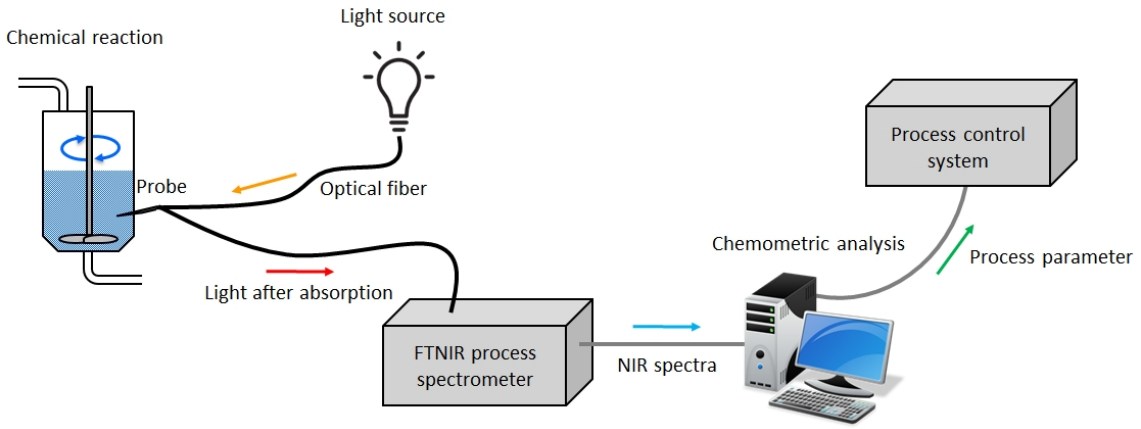

Figure 1: Scheme of the data acquisition process and its components in inter-communication with the chemical reactor

which is robust with respect to temperature changes [19] and uses a semiconductor reference laser. The resolution and the measurement rate of the instrument are configurable. The programmable logic controller, used for reading data acquired by sensors and controlling actuators, employs a field-bus interface using the CANopen protocol. Unlike conventional PLCs it supports large data structures like infrared spectra. Furthermore, it provides mechanisms for storing measurement data permanently to the 'process control system'.

A schematic overview of the data acquisition process and its interconnecting components at the process in the loop and in communication with the chemical reactor is visualized in Figure 1. A halogen light source provides infra-red light to a Hellma transmission probe (optical path length: 1 mm) immersed directly into the resin inside the reactor. Despite accumulation of (solid) MF resin on the immersion probe at low temperatures, polymerisation can be adequately monitored during the critical phase of the process (i.e. towards the end of the reaction). An optical fiber (length: 2×0.25 m), using quartz glass as fiber material (fiber type: AS600/660IRAN), is connected to the light source (halogen lamp) connecting the probe with the spectrometer. Its maximal and minimal immersion depth is 690 mm and 30 mm, respectively. The outer diameter is 25 mm and the pipe length of the probe is 700 mm. The temperature range of the product lies between 5°C and 260°C and the pressure range between 1 and 35 bar. The spectral range of the current setup expands from

820 nm to 2857 nm yielding spectra comprising a total of 2753 variables. The final spectra are calculated as the average of 80 single shot interferograms, acquired during *ca.* 8 seconds.

During the condensation process, FT-NIR spectra and several process parameters (e.g. temperature and pressure) are continuously collected from the process control system in real-time via CANopen and permanently stored in a separate data server. Three consecutive spectra are recorded within a few seconds and averaged in order to reduce the noise effect in single measurements. The averaged spectra are then used for prediction and (eventual) model updates if necessary. The predicted values (i.e. turbidity points) are provided in real-time to the process control system. The optimal endpoint of a batch is automatically determined based on prediction of the DP without human intervention. In a second stage, the resin solution is cured using defined heat and pressure yielding highly cross-linked MF resins.

3. Theory

3.1. Notation

Scalars are denoted by italic symbols (e.g. x or V). Lowercase and uppercase boldface symbols denote column vectors ($\mathbf{t}$) and matrices ($\mathbf{X}$), respectively. Unless explicitly stated, vectors are column vectors and superscripted symbols ^{T} and $^{-1}$ indicate the transpose and inverse, respectively of a vector or matrix. Vertical concatenation of column vectors is indicated using comma notation (e.g. $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_A]$). $\mathbb{E}[x]$ denotes the expected value of a random variable x.

3.2. Chemometric Models

3.2.1. PLS and ensemble PLS

Partial least squares (PLS) is aimed at finding a set of orthogonal direction vectors $\mathbf{w}_a \in \mathbb{R}^{K\times 1}$ for $a \in \{1, \ldots, A\}$ such that projection of a matrix $\mathbf{X} \in \mathbb{R}^{N\times K}$ (of N samples and K covariates) on $\mathbf{w}_a$ – termed scores $\mathbf{t}_a$ – have maximal covariance with some target vector $\mathbf{y} \in \mathbb{R}^{N\times 1}$.

Formally,

$$\begin{aligned} &\mathbf{w}_a = \arg\max_{\mathbf{w}} \; cov(\mathbf{X}\mathbf{w}, \mathbf{y}) \\ &\text{s.t. } \; \|\mathbf{w}_i\| = 1 \text{ and } \mathbf{w}_i^T \mathbf{X}^T \mathbf{X} \mathbf{w}_i = 0 \quad \text{for } 1 \leq i \leq a. \end{aligned} \tag{1}$$

In the second step of PLS, $\mathbf{y}$ is regressed on these new latent variables $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_A]$. Since usually $A << K$ sufficiently captures most of the variance in $\mathbf{y}$ and the scores $\mathbf{t}_a$ are mutually orthogonal, $\mathbf{T}^T\mathbf{T}$ has full (column) rank and can thus be safely inverted circumventing the usual pitfall of least squares regression when coping with fat data matrices and/or highly correlated variables (which is usually the case for spectroscopic data). For a comprehensive introduction on PLS, we refer to [20] and for its application within an ensemble modelling scheme to [21]. We will exploit the latter by generating bootstrap samples from the calibration data [22]. In particular, for a given data set $\mathbf{X}$ containing N samples, we randomly draw N samples with replacement and store them in a separate matrix $\mathbf{X}_1$ (termed one bag) — this procedure is repeated B times, yielding B bags $\mathbf{X}_1, ..., \mathbf{X}_B$ each containing N samples. For each of the bags $\mathbf{X}_1, ..., \mathbf{X}_B$, a separate PLS regression (PLSR) model is trained. During the prediction phase, the average prediction over all models is calulated to obtain the prediction $\hat{y}_i$ on the i-th test sample,i.e.

$$\hat{y}_i = \frac{1}{B} \sum_{j=1}^{B} \hat{y}_i^{(j)}. \tag{2}$$

3.2.2. FLEXFIS-PLS (non-linear version of PLS)

We extend PLS to a non-linear version in order to be able to model any intrinsic non-linearity between latent features and reference measurements (i.e. turbidity points). In order to circumvent large-scale data matrices and/or extra (kernel) parameters as used in non-linear variants of PLS proposed in literature so far (such as in kernel-PLS [23]) and PLS variants employing non-linear modelling of the inner relationship between $\mathbf{X}$ and $\mathbf{y}$ [24] [25], we employ fuzzyfication of the latent variable space as previously described in [15] and [26], which allows capturing local data structures over the feature space – as exemplarily shown in Figure 2 for a simple two-dimensional example from a real-world process. Obviously, partitioning the data into two local regions and modelling

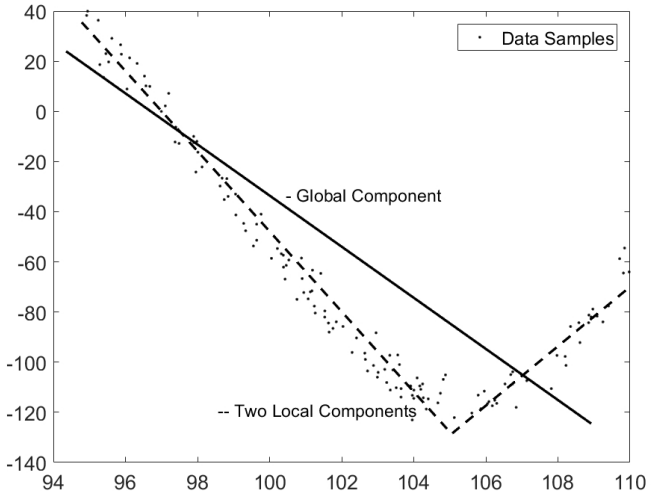

Figure 2: A two-dimensional example of local data clouds showing two components/trends (as dashed lines), which cannot be sufficiently resolved by a global, linear model (solid line).

the relationship between x and y for each region separately (dotted lines) is more appropriate than a global, linear model (solid line). The idea of our fuzzy modelling component is to obtain partial piece-wise linear predictors l_i for $i = 1, ..., C$ local regions of the latent variable space:

$$l_i(\mathbf{t}) = \beta_{i0} + \beta_{i1} t_1 + \beta_{i2} t_2 + ... + \beta_{iA} t_A, \tag{3}$$

amalgamated with multi-dimensional fuzzy rules defined through a multi-dimensional Gaussian kernel (implying ellipsoidal contours) to achieve smooth model outputs:

$$\hat{y} = \sum_{i=1}^{C} l_i(\mathbf{t}) \Psi_i(\mathbf{t}) = \sum_{i=1}^{C} l_i(\mathbf{t}) \frac{\mu_i(\mathbf{t})}{\sum_{j=1}^{C} \mu_j(\mathbf{t})} \qquad \mu_i(\mathbf{t}) = exp(-\frac{1}{2}(\mathbf{t} - \mathbf{c}_i)^T \mathbf{\Sigma}_i^{-1} (\mathbf{t} - \mathbf{c}_i)). \tag{4}$$

$\mathbf{t}$ is the $A \times 1$ scores vector of a projected test sample, $\mathbf{c}_i$ is the centroid of the i-th rule (i.e. the i-th local region of the latent variable space) and $\mathbf{\Sigma}_i^{-1}$ the corresponding local covariance matrix. This has some similarities with locally weighted regression [27], but produces a global, more interpretable model. For eliciting the optimal number of fuzzy rules (according to the non-linearity in the input-output relationship) and for synchronous estimation of the parameters in the antecedent

(i.e. $\mathbf{c}_i, \mathbf{\Sigma}_i^{-1}$) and consequent (i.e. $\boldsymbol{\beta}$) space, we employ a generalized evolving fuzzy systems learning engine termed *Gen-Smart-EFS* [28]. A special case comes up when off-diagonal elements of the inverse covariance matrix $\mathbf{\Sigma^{-1}}$ are all set to 0 (ignoring the covariances between the inputs). In this case, axis-parallel fuzzy rules are triggered and the steps in the itemization above end up in the classical flexible fuzzy inference systems (FLEXFIS) approach [29], which we thus term *FLEXFIS-PLS*.

3.3. Model Adaptation

3.3.1. Update of PLS Models and Ensembles

For (ensemble) PLS models we propose a forgetting based weighting scheme in order to give new samples more emphasis during model re-calibration and outweigh older samples in the data stream. This leads to an increased flexibility when adapting to new 'drifted states'. In particular, we define a weighting function

$$\omega(i) = \lambda^{N-i}, \tag{5}$$

with λ being a forgetting factor typically set to a value between 0.9 and 1 (default 0.95), N is the number of labeled spectra in the actual calibration set (i.e. samples for which reference measurements are available) and $i \in \{1, ..., N\}$ denotes time increments (from old to new). We normalize $\omega(i)$ such that $\sum_1^N \omega_i = N$ and define a $N \times N$ diagonal matrix $\mathbf{G}$ with the i-th element being $g_{i,i} = \omega_i$. Finally, we solve the weighted PLS problem by replacing $\mathbf{X}$ with $\mathbf{X}_G = \mathbf{GX}$ in (1) when a new reference measurement becomes available.

3.3.2. Active Learning

For supervised model updates, a sample selection strategy becomes indispensable in order to keep measurement costs low. We here use violation of the critical limits of the well established Hotelling's T^2 and Q-Residual statistics as criterion to request reference measurements of incoming test samples as described previously in [12]. Hotelling's T^2 is proportional to the distance of a

projected sample to the training data centroid in the latent variable space. Per definition

$$T^2 = \sum_{j=1}^{A} \left(\frac{t_j}{\lambda_j} \right)^2, \tag{6}$$

where A denotes the total number of latent variables considered in the model and t_j denotes the score of a test spectrum when projected on the j-th loading vector $\mathbf{p}_j$ with eigenvalue λ_j. Critical limits for T^2 are calculated by exploting the relationship [30]

$$T^2 \sim \frac{A(K-1)}{K-A} F_{(A,K-A)}, \tag{7}$$

where A and K denote the number of LVs and number of original variables, respectively. $F(A, K-A)$ is the Fischer-Snedecor distribution with A and $K-A$ degrees of freedom in the numerator and denominator, respectively. In contrast to Hotelling's T^2, which only considers the distance in the projected space, Q-Residuals are a measure of how well a test datum is represented by this subspace. Geometrically, they can be regarded as (orthogonal) distance between the datum in the original (high dimensional) space and its projection on A-dimensional latent feature space. For the i-th test datum

$$Q_i = \|\mathbf{x}_i - \mathbf{x}_i \mathbf{P}_A \mathbf{P}_A^T\|_F^2. \tag{8}$$

$\mathbf{P}_A \in \mathbb{R}^{KxA}$ is the loadings matrix that projects x_i to the A dimensional latent feature space. Similar to Hotelling's T^2, upper confidence bounds on Q-Residuals are often used in process monitoring in order to check whether spectra from incoming samples are well described by the model. We here used the Jackson-Mudholkar approximation [30]

$$Q_\alpha = \Phi_1 \left[\frac{c_\alpha \sqrt{2\Phi_2 h_0^2}}{\Phi_1} + 1 + \frac{\Phi_2 h_0 (h_0 - 1)}{\Phi_1^2} \right]^{1/h_0} \tag{9}$$

with

$$\Phi_i = \sum_{j=A+1}^{M} \lambda_j^z \text{ , for } z = \{1, 2, 3\} \text{ , } M = min(N, K) \tag{10}$$

and

$$h_0 = 1 - \frac{2\Phi_1 \Phi_3}{3\Phi_2^2}. \tag{11}$$

c_α is the standard normal deviate corresponding to the upper (1-α) quantile and α is typically set to 0.95 or 0.99. As a lack of fit statistic, Q-Residuals tend to exceed these confidence limits if test data contain unmodelled information, which is usually indicative of a change in covariance structure (e.g. due to the presence of a new interferent). Combined together, a reference measurement for the i-th test sample is requested if

$$T_i^2 > T_\alpha^2 \text{ or } Q_i > Q_\alpha, \tag{12}$$

with T_α^2 and Q_α denoting the critical limits for Hotelling's T^2 and Q-Residuals calculated according to (7) and (9). α was systematically varied in order to investigate the relationship between critical limits, number of selected AL samples and resulting prediction error, (see Section 4.3). For ensemble PLS, a reference measurement was requested if either limit in (12) was violated for at least one of the models in the committee.

3.3.3. FLEXFIS-PLS Model Update in Unsupervised Mode

Unsupervised adaptation of FLEXFIS-PLS models is carried out by first determining the closest rule from a test spectrum in terms of the Mahalanobis distance to the corresponding rule's center in the latent variable space. The position of the winning rule

$$\mathbf{c}_{win} = \underset{i=1,\ldots,C}{\arg\min}\, d(\mathbf{t}, \mathbf{c}_i) \tag{13}$$

is then updated according to

$$\mathbf{c}_{win}(N_{win}+1) = \mathbf{c}_{win}(N_{win}) + \frac{\omega(\mathbf{t})}{\eta_{win}}(\mathbf{t} - \mathbf{c}_{\mathbf{win}}(N_{win})). \tag{14}$$

$\mathbf{t}$ denotes the scores vector obtained by projection (i.e. $\mathbf{t} = \mathbf{xW}$), N_{win} denotes the support of the winning rule (i.e. the number of calibration samples contained in it), $\omega(\mathbf{t})$ is the weight assigned to the current sample according to (5) and η_{win} is the learning gain for the winning rule (i.e. $\frac{1}{N_{win}}$). Likewise, the spread of the winning rule (i.e. the variance σ_{win}^2 in each direction), is adapted through recursive variance update including rank-1 modification for increased stability [31]. where

the new contribution of the current sample to the variance, i.e. $(\mathbf{t} - \mathbf{c}_i(N_{win} + 1))^2$, is weighted by $\omega(\mathbf{t})$:

$$\sigma^2_{win}(N_{win} + 1) = \frac{N_{win}}{N_{win} + 1}\sigma^2_{win}(N_{win}) + \Delta c^2_{win} + \frac{\omega(\mathbf{t})}{N_{win} + 1}(c_{win}(N_{win} + 1) - \mathbf{t})^2 \tag{15}$$

with $\Delta c^2_{win} = (c_{win}(N_{win}) - c_{win}(N_{win} + 1))^2$, i.e. the difference between the updated and the old position of the cluster center (rank-1 modification [31]). In contrast to the original, fully (unweighted) supervised adaptation of FLEXFIS-PLS [15], the consequent parameters (i.e. the coefficients for the local linear predictors) are kept fixed (once estimated from an initial batch training set). Hence, only adaptation of the positions and spreads of the rule contours (defined though multivariate normal distributions) is performed in order to account for changes in the input space.

3.4. Drift Detection

3.4.1. Test of Page and Hinkley

The test of Page-Hinkley (PH) is a cumulative sum based change detector of some normally distributed quality indicator θ [32] [33]. For a test datum, sampled at time point T, the PH statistic is defined as:

$$\mathrm{PH}_T = S_T - \min_{0 \le t \le T} S_t \tag{16}$$

with

$$S_T = \sum_{t=1}^{T} (\theta_t - \bar{\theta}_{T-1} - \delta). \tag{17}$$

$\bar{\theta}_{T-1}$ is the average of all θ's up to time point $T - 1$ and δ is the magnitude of the acceptable change in θ (usually set close to 0). A drift is detected if $\mathrm{PH}_T > \lambda$. Lughofer *et al.* [16] adapted the test in the context of stream mining by introducing a fading factor $\alpha \in [0, 1]$ to outweigh older samples in the stream yielding

$$S_T = (\theta_T - \bar{\theta}_{T-1} - \delta) + \alpha S_{T-1}, \tag{18}$$

This faded version of the PH test is particularly useful if several drifts are present in a data stream [16]. In order to become independent from setting the threshold λ to a fixed value, we further

propose a modification involving incremental updates of mean and standard deviation of the PH statistic according to

$$\overline{\mathrm{PH}}_T = T * \overline{\mathrm{PH}}_{T-1} + \frac{1}{T+1}\mathrm{PH}_T \tag{19}$$

and

$$\sigma_{\mathrm{PH}_T} = (\overline{\mathrm{PH}}_T - \overline{\mathrm{PH}}_{T-1})^2 + \sqrt{\frac{T}{T+1} * \sigma^2_{\mathrm{PH}_{T-1}} + \frac{1}{T+1}(\overline{\mathrm{PH}}_T - \mathrm{PH}_T)^2}, \tag{20}$$

respectively and defining (pseudo-) statistical control limits at $\overline{\mathrm{PH}}_t \pm 3\sigma_{\mathrm{PH}_t}$ at each instance $t = \{1, \ldots, T\}$. Violation of these limits indicates that a drift has occurred, i.e. whenever either

$$\mathrm{PH}_T \geq \overline{\mathrm{PH}}_{T-1} + 3\sigma_{\mathrm{PH}_{T-1}} \tag{21}$$

in case that θ is increasing or

$$\mathrm{PH}_T \leq \overline{\mathrm{PH}}_{T-1} - 3\sigma_{\mathrm{PH}_{T-1}}, \tag{22}$$

otherwise.

3.4.2. Committee Disagreement (CD)

In order to assess the quality θ of a prediction from a chemometric model we propose the committee disagreement degree (CD), which is computed as the variance among predictions from individual models of an ensemble trained on different subsets of the calibration set. For the i-th test spectrum

$$\mathrm{CD}(x_i) = \frac{1}{B}\sum_{j=1}^{B}(V^{(j)}(x_i) - \bar{V}(x_i))^2, \tag{23}$$

where $V^{(j)}(x_i)$ denotes the prediction from the j-th model and $\bar{V}(x_i)$ is the average prediction on x_i among the B models in the committee. It can be shown that $\mathbb{E}[\mathrm{CD}]$ is a lower bound for the squared generalization error of the committee and an increasing CD over time, on average, thus indicates that the error associated with model predictions is increasing [34]. Similar to Hotelling's T^2 and Q-Residuals, no target values are required to compute the CD, which is in contrast to other supervised

approaches for drift detection (based on model errors etc.) [35] [36] [37]. However, in contrast to these more traditional approaches to anomaly detection in chemometrics, the CD is based on actual predictions rather than on mere outlyingness with respect to the calibration samples or the latent variable subspace. This latter point is important since multivariate calibration models often tend to be capable of extrapolating considerably beyond the training samples and consequently may deliver accurate predictions despite significant violations of statistical control limits [6].

3.4.3. Other quality metrics for assessing the quality of predictions

Sample-Wise Prediction Uncertainty. In order to estimate sample-wise prediction uncertainty, we here use the approximation proposed by Faber and Bro [38], i.e.

$$\sigma_{\mathrm{PE}} \approx \sqrt{(1 + h_{test}) * \mathrm{MSEC} - V_{\Delta y}}, \tag{24}$$

where $h_{test} = \mathbf{t}_{test}^{T}(\mathbf{T}^{T}\mathbf{T})^{-1}\mathbf{t}_{test}$ is the leverage of a test spectrum, $\mathbf{t}_{test} \in \mathbb{R}^{Ax1}$ denotes the scores obtained by projection on the A latent variables retained in the model and $\mathbf{T} \in \mathbb{R}^{NxA}$ is the scores matrix from the training samples. The mean squared error of calibration (MSEC) is obtained in the usual way and $V_{\Delta y}$ denotes the variance (uncertainty) of the reference measurements in the training set. The latter was set to 0 in all experiments since we were interested in how uncertainty changes (increases) when spectra become affected by potential drifts rather than estimating the uncertainty itself.

Hotelling's T^2 and Q-Residuals. Although T^2 and Q-Residuals (Section 3.3.2) are no direct measures for the reliability of predictions from latent variable models they are widely used in chemometrics as a guide for practitioners to decide if normal operation conditions are violated. However, as stated by Wise *et al.* [6], multivariate calibration models are often capable to considerably extrapolate beyond the widely used 95% and 99% confidence limits, which is particularly true for the former. In case of high Q-residuals - usually indicative for the presence of a new source of variation that has not been accounted for in the calibration step (e.g. due to the presence of a new interferent) — prediction accuracy is affected only in case that the net analyte signal is affected.

Table 1: Dataset characteristics.

Scenario	# Variables	# Samples		# Drifts
		Cal	Test	
1	346	7521	1251	1
2	346	1442	129	3
3	346	2334	720	2

4. Experimental Setup

4.1. Dataset Characteristics and Data pre-treatment

Three drift effects were observed during the observation horizon, which are covered by three datasets further referred to as scenarios 1-3 (Table 1). All spectra were smoothed using the Savitzky-Golay algorithm (15 smoothing points, 0-*th* order polynomial) and subsequently normalized by the standard normal variate (SNV) normalization method applied to the entire spectral range comprising at total of 2753 variables. All models were calibrated using a combination of the following spectral ranges: 1599 – 1824 nm and 1434 – 1516 nm (Figure 3). These two spectral ranges include the carbon-hydrogen and especially the oxygen-hydrogen and water bands which typically change during the condensation process and are thus most suitable for process monitoring. The selection of the relevant spectral ranges was performed manually based on chemical knowledge of experts at Metadynea Austria GmbH and RECENDT GmbH.

4.2. Model Parametrization

4.2.1. PLS

PLS models were fitted using PLS-Toolbox 8.2.1 (Eigenvector Research, Manson, WA, USA) and Matlab 2015a (Mathworks, Natick, MA, USA). The optimal complexity of PLS models was determined by 10-fold cross-validation (CV). To this end, the number of latent variables was varied between 1 and 15 and the root mean squared error of cross-validation (RMSECV) calculated in the usual way. Minimization of the CV error w.r.t the number of LVs often yields PLS models that are

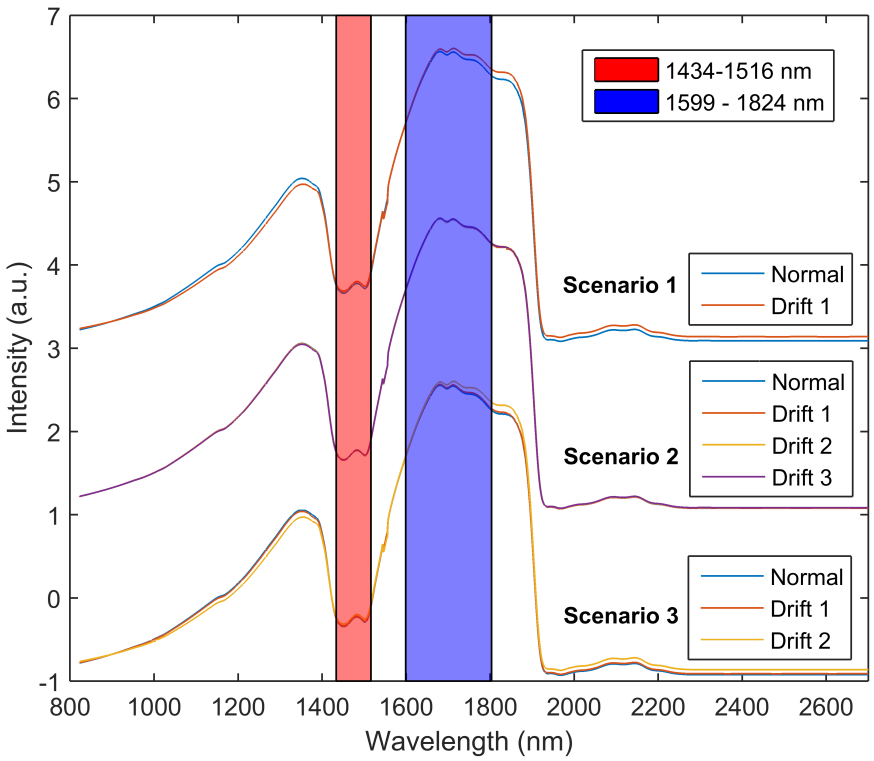

Figure 3: Representative FT-NIR spectra of Melamine Formaldehyde (MF) resin. Wavelength regions used for modelling the turbidity point are indicated. An average spectrum representing the normal condition and average spectra of drifted states are shown for each of the 3 investigated scenarios.

too complex. We therefore choose the optimal number of LVs by minimizing a penalized CV error criterion

$$\text{RMSECV}_{pen} = \text{RMSECV} \times \exp(0.05 \times \#\text{LVs}), \tag{25}$$

which includes a penalty for the number of latent variables (Figure 4). Ensemble PLS (ePLS) models were established by taking bootstrap samples from the pool of calibration data. The number of individual PLS models (i.e. committee members) was set to 20 in all experiments to achieve a fair comparison — based on several initial experiments, it turned out that this number is pretty insensitive for the results, anyway.

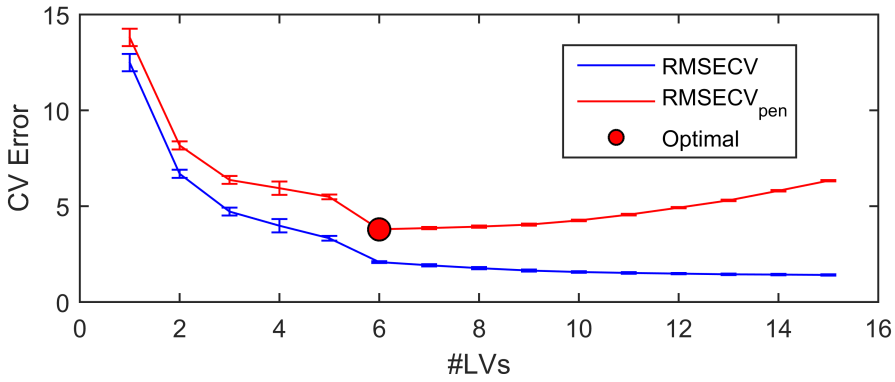

Figure 4: Model selection. The optimal model complexity in terms of number of latent variables (LV) is determined by minimizing (25) in order to penalize for inclusion of too many LVs.

4.2.2. *FLEXFIS-PLS*

In addition to the number of LVs, the number of fuzzy rules controlling the degree of non-linearity in the input-output relationship has to be tuned for FLEXFIS-PLS. To this end we set up the search grid for the number of LVs and fuzzy rules in the range between 1 and 15 for the former and 1-10 for the latter. Each parameter combination was evaluated by 10-fold CV and the resulting CV error penalized for complexity yielding a penalized CV error criterion (as an extension of the criterion in (25)):

$$\text{RMSECV}_{pen} = \text{RMSECV} \times \exp[0.05(\#\text{LVs} + \#\text{Fuzzy Rules})], \tag{26}$$

which was minimized in order to obtain the optimal parameter combination for the final model. The value of 0.05 in the exponential term is a default value we successfully used for various other applications and (spectroscopic) data in the past.

4.3. *Method Evaluation*

Performance Comparison. The performance of calibration models with and without model adaptation was compared by means of the global root mean squared error of prediction

$$\text{RMSEP}_{glob} = \sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}(\hat{y}_i - y_i)^2}, \tag{27}$$

where $\hat{y}_i$ and y_i denote predicted and measured turbidity points and N_t is the number of test samples used for evaluation. Since the analytical goal in melamine resin production is to accurately determine when condensation should be stopped, we also assessed the RMSEP locally in the region $y = [20, 40]$ close to the end point of the reaction, i.e.

$$\mathrm{RMSEP}_{loc} = \sqrt{\frac{1}{N_{[20,40]}} \sum_{y \in [20,40]} (\hat{y}_i - y_i)^2}. \tag{28}$$

In order to systematically examine active learning, significance levels α in (12) were varied within the values $[0.999, 0.99, 0.95, 0.9, 0.75, 0.5]$ in order to change the frequency of reference measurements and model updates. In addition, active sample selection was compared to passive selection in terms of RMSEP improvements over static models. For the latter, equidistant sampling was used at the same rate as in active learning at a given α to allow direct comparison.

5. Results

5.1. Drift Detection

To study the merit of incrementally adapting control limits on the Page-Hinkley (PH) statistic, different model diagnostic measures including prediction uncertainty, Hotelling's T^2, Q-Residuals and committee disagreement (CD) were tracked over time in 3 different drift scenarios (Figure 5). Since all of these quantities tend to increase as predictions become biased, only violation of the upper control limit on the PH statistic was considered for drift detection; i.e., only equation (21) was used and shown as black-lined threshold in Figure 5. The top row in figure 5 shows the evolution of the residuals (i.e. absolute difference between measured and predicted turbidity points) produced by the corresponding chemometric models (without model adaptation). In the first scenario, considerable prediction bias is introduced around sample 650 with predictions for downstream samples becoming considerably shifted. In the second and third scenario, prediction bias is transient and the models deliver accurate predictions in between. In addition, predictions start drifting severely towards the end of the recordings for the latter. Analysis of the control charts based on the PH statistic revealed that prediction uncertainty ($\hat{\sigma}_{PE}$) and Hotelling's T^2 failed to report timely on the

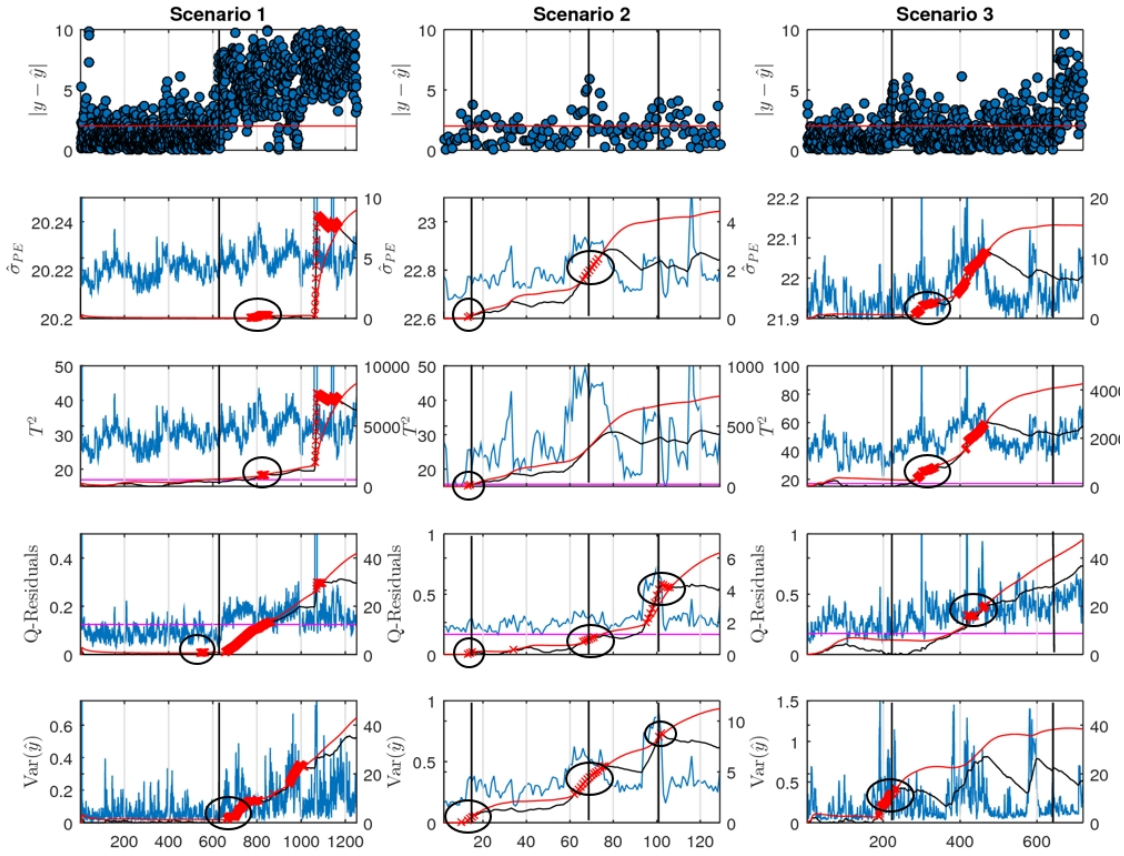

Figure 5: Drift detection by faded Page-Hinkley test using incrementally evolving control limits. Top row: Absolute difference between observed and predicted turbidity point temperatures for three independent test scenarios using static (SoA) calibration models. The horizontal line indicates the desired analytical accuracy of the model. Subsequent rows show evolution of corresponding model diagnostic measures for each scenario in the following order: $\hat{\sigma}_{PE}$: prediction uncertainty (24), T^2: Hotelling's T^2(6), Q-residuals (8), Var($\hat{y}$): Committee disagreement (23). Blue lines indicate the raw data, evolution of the Page-Hinkley statistic and the corresponding upper threshold according to (21) are shown as black and red lines, respectively. Red crosses mark violations of the upper control limit by the the PH statistic. Solid ellipsoids mark those violations closest to real drifts, which are indicated by vertical lines and which have been determined by the experts from the company partner.

drift seen in scenario 1, whereas upper adaptive control limits were violated around the time point when the drift occurred for the Q-residuals and the committee disagreement. However, the former gave rise to an alarm fairly before the drift actually occurs, which is considered a false alarm. In

Table 2: Comparison of the performance of different drift detection methods based on four indicators, including our proposed CD (committee disagreement); bold font indicate best results over the four variants.

Indicator	Detection Capability	Sign. Delays (> 50 samples)	# of False Alarms
σ_{PE}	4 out of 6	2 times	2
T^2	3 out of 6	2 times	2
Q	4 out of 6	**0 times**	3
CD (ours)	**5 out of 6**	**0 times**	**1**

scenario 2, the PH statistic successfully indicated occurrence of a drift around sample No. 12 for all investigated diagnostic measures. In addition, the drop in accuracy around sample 70 and 100 was successfully reported by either Q-Residuals and the CD, but not by Hotelling's T^2. The behavior of the diagnostic measures was most divergent for scenario 3 with the CD successfully reporting on the increase in prediction error around sample 200. However, all variants failed to detect the second drift towards the end of the stream in this scenario. In contrast, the PH statistic computed from the Q-Residuals did not indicate unusual behavior of the model around sample 200 but violation of the control limit was given around sample 400, which however must be regarded as a false alarm. In addition to violations around sample 400, Hotelling's T^2 and $\hat{\sigma}_{PE}$ displayed violations around 250 – slightly delayed with respect to the high errors seen around sample 200. Finally, we observed that the PH statistic associated with CD (and to some extent also with the Q-Residuals) *correctly* decreased as predictive accuracy increased again in case of the transient effects causing prediction bias in scenario 2 and 3. In case of scenario 2, the third drift was solely detected by the CD and the Q-Residuals. The overall performance over all six drift cases in the three streams is summarized in Table 2, reflecting more clearly the out-performance of SoA indicators by our proposed committee disagreement variant: the columns denote i) the detection capability in terms of the number of drifts that were correctly detected, ii) the number of significant delays (> 50 samples) between drift and detection and iii) the number of false alarms that were triggered by the corresponding model diagnostic measure. These numbers are calculated based on the fact that

the experts saw a significant delay in detecting a drift correctly in case of more than 25 samples. Due to an implicit uncertainty of the experts about the exact location of a drift, there has been also respected a tolerance of 25 samples before and after the vertical lines (indicating the most likely drift positions defined by experts) when evaluating the drift detection performance. This means that no false alarm is counted when a drift is detected up to 25 samples before, and no significant delay is counted when a drift is detected up to 50 samples (25 + 25 — as indicated in the first row of Table 2) after the expert-defined drift positions.

5.2. Active Learning (AL)

In order to guide adaptation of chemometric models when a drift has been detected, we first investigated a previously described AL strategy that builds upon the idea to request reference measurements for samples (spectra) that violate critical limits of Hotelling's T^2 and/or Q-residuals. As soon as the reference information is available, the calibration set is extended by this sample and the model automatically re-calibrated using the techniques described in section 3.3. We here further extended this approach by incorporating gradual forgetting of older samples and by placing more emphasis on newer samples, which is in contrast to our previous sliding window approach that eventually discards samples encoding important process variability. The automatic model update engine is activated upon the first violation of the incrementally adapting threshold on the PH statistic associated with the CD as it was the overall best measure for drift detection over all 3 scenarios. Critical limits for Hotelling's T^2 and Q-Residuals were systematically varied in order to control the number of violations and thus the frequency at which reference measurement should be performed. Figure 6 compares model adaptation based on active vs. passive (i.e. random) selection of samples in the 3 scenarios. We found that active sample selection performed favorably over random selection in scenario 1. Thereby, accuracy could be significantly increased from RMSEP = 3.56 corresponding to the reference model (without adaptation) to an acceptable level (i.e. RMSEP between 1.9 and 2.5) when requesting reference measurements at a rate of $5 - 10\%$ corresponding to $\approx 30 - 60$ samples over the entire stream (Table 3). In contrast, active selection did not yield consistently better results than passive selection in scenarios 2 and 3, which can be attributed to the

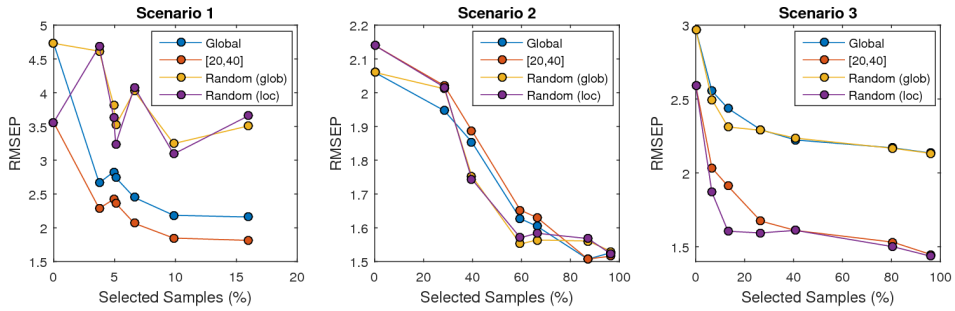

Figure 6: Active learning vs. random selection of samples for ensemble PLS. Global and local (i.e. in the turbidity point range of [20,40]) root mean squared errors of prediction are shown for sample selection by active learning and equidistant, random selection at the same sampling frequencies.

transient nature of the drifts in these scenarios. While the AL engine runs into danger of triggering re-calibration based on outlying samples leading to deterioration of predictive accuracy, equidistant (random) selection can be more robust with respect to putative outliers or transient drifts. This effect was most pronounced when the fraction of samples selected for model adaptation was small and flattens out as more re-calibration samples were included. Yet overall, accuracy could be kept within an acceptable range in scenarios 2 and 3 irrespective of how re-calibration samples were selected.

5.3. Unsupervised Model Adaptation

We recently introduced a fuzzy systems version of PLS (FLEXFIS-PLS) that performs partitioning of the latent variable space, similar to locally weighted regression, in order to model non-linear relationships between inputs and targets. Notably, the FLEXFIS-PLS architecture allows unsupervised adaptation of trained models in single-pass mode (i.e. without the need to re-evaluate the entire calibration set). To this end, the center and spread of the local region closest to the current sample is incrementally updated towards this sample. This in turn gradually changes the membership degree function in Equation (4) and thus the weighting of the local linear predictors during inference.

We first investigated if the static version of non-linear FLEXFIS-PLS can outperform ordinary,

linear ensemble PLS on the three scenarios (without adaptation). On the one hand, we found that increasing the number of fuzzy rules beyond 1 did in general not increase accuracy of FLEXFIS-PLS models (results not shown) indicating that the relationship between the LVs and the turbidity point is highly linear. On the other hand ePLS and static FLEXFIS-PLS – which correspond to ordinary PLS if the number of fuzzy rules is 1 – performed comparably except for scenario 1 where ePLS significantly outperformed static FLEXFIS-PLS. This finding is in line with an improved bias-variance trade-off resulting from bagging as discussed in detail in [39]. We next investigated the ability of the FLEXFIS-PLS architecture to adapt to the drifts through single-pass updates of the models' antecedents when the number of fuzzy rules was > 1 (Figure 7), termed as unsupervised adaptation in Table 3 below. Similar to active learning, adaptation was initiated by the first detection of a drift. However, in contrast to AL, all subsequent samples were used for model adaptation without restriction. Overall, we found significant improvements in terms of the RMSEP in scenario 2 and 3, while only the global error could be improved with respect to the reference model (without adaptation) in scenario 1. This latter finding can be explained by the abundance of low turbidity point samples (for which $y << 20$) in the first dataset due to the fact that process monitoring was carried out manually after occurrence of the drift. We further found that the ability to adapt to the drifts was strongly dependent on the number of fuzzy rules employed, which corresponds to the non-linearity degree of the models. In scenario 2, two fuzzy rules yielded the best performance, while 8 rules were optimal in scenario 3 (Table 3). Increasing the non-linearity degree beyond the optimal value led to rapid increase of prediction errors, which can be attributed to overfitting.

5.4. Discussion and Overall Performance

We here applied the Page-Hinkley test to detect changes in different model diagnostic quantities associated with chemometric models in order to determine the time point when initially deployed models should be adapted. In our extension, control limits are derived autonomously by incremental estimation of the mean and standard deviation of the PH statistic, which proved successful for reliable detection of abrupt changes in prediction accuracy irrespective of the scale of the underly-

Table 3: Results summary. The (average) number of latent variables retained in the model, number of fuzzy rules employed for unsupervised model adaptation, local (i.e. in the turbidity point range of [20,40]) and global RMSEP for each method are shown. For active learning, RMSEP values refer to a frequency of ≈ 10% reference measurements used for model adaptation.

	Method	# LVs (# Fuzzy Rules)	RMSEP local	RMSEP global
Scenario 1	Static ePLS	9	3.56	4.73
	Unsupervised Adaptation	9 (1)	3.7	4.8
	Active Learning (≈ 10%)	9	1.85	2.19
Scenario 2	Static ePLS	8	2.06	2.14
	Unsupervised Adaptation	8 (2)	2.01	2.10
	Active Learning (≈ 10%)	8	2.07	2.05
Scenario 3	Static ePLS	8	2.62	3.00
	Unsupervised Adaptation	8 (8)	2.16	2.64
	Active Learning (≈ 10%)	8	1.6	2.0

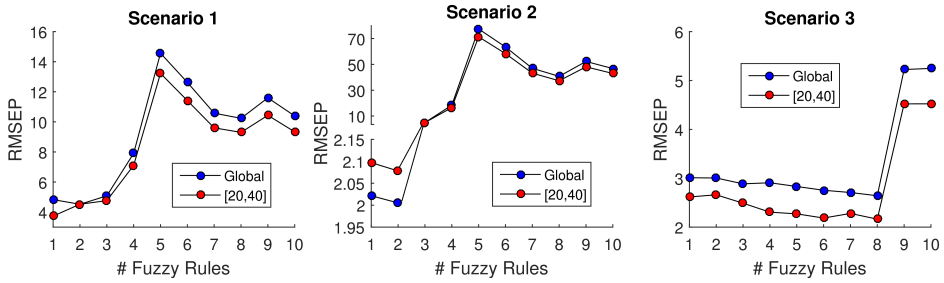

Figure 7: Unsupervised model adaptation by FLEXFIS-PLS. The change of global and local (i.e. in the turbidity temperature range of [20,40]) RMSEP vs. number of fuzzy rules is shown for scenarios 1-3. Data points where # Fuzzy Rules = 1 indicate the accuracy of FLEXFIS-PLS without adaption.

ing diagnostic measure. This latter point is important since in real world applications, diagnostic measures often tend to violate (rigid) statistical thresholds without implications in terms of prediction bias. For instance, Hotelling's T^2 was trending above the 95%-CI in all 3 scenarios most of the time while predictions were (mostly) fine. In addition, the most common diagnostic measures in chemometrics (i.e. T^2 and Q-Residuals) relate to some sort of similiarity between test datum and calibration samples without taking into account the inference process. To this end we proposed the committee disagreement (CD) as a diagnostic measure, which takes this aspect into consideration. Notably, we found that the CD was the only quantity for which our drift detector gave an alarm when prediction bias significantly increased for the first time in all 3 investigated scenarios underpinning the importance of this aspect. Once a drift has been confirmed, active learning based on violations of Hotelling's T^2 and Q-residuals proved successful to keep the overall accuracy of chemometric models within the analytical goal while at the same time minimizing the number of reference measurements required for adaptation. AL proved particularly successful when predictions from initial models became persistently biased (scenario 1). In case of transient drifts on the other hand (scenarios 2 and 3), AL could in fact not outperform random selection when using the same low number of samples ($< 10\%$) for model adaptation. However, these results do not generally lead us to conclude that AL is not suitable in case that a drift is only transient: The fact that adaptation based on AL did not significantly increase prediction bias compared to random sample

selection actually underpins the robustness and flexibility of AL in combination with automated model re-calibration.

Regarding unsupervised model adaptation of the fuzzy rule-bases in FLEXFIS-PLS, our results clearly demonstrate significant improvements in all scenarios (compared to static FLEXFIS-PLS models). However, model adaptation could not appropriately compensate for the severe drift in scenario 1, which can be attributed to the fact that a strong change in covariance structure occurred in this scenario (indicated by the sudden increase of the Q-residuals around sample 660). In such case, adaptation of the latent variable space based on additional (reference) measurements becomes indispensable. For the lighter drifts (scenarios 2 and 3), unsupervised model adaptation could significantly reduce the RMSEP of static models provided that model complexity (# Fuzzy Rules) was appropriate. Yet, model selection by traditional cross-validation on the calibration data is of little use for the determination of the optimal complexity since optimality largely depends on the particular drift situation, which is not known beforehand. Future work on unsupervised model adaptation should therefore address this dilemma in more detail.

Finally, in Table 4 we compare the overall performance of the calibration model maintenance strategies in terms of their cost-benefit tradeoff including results based on sliding window SW-PLS with and without active learning (but without explicit drift detection) as applied and compared before in [40] for exactly the same application scenario (the same production site). Notably, reference sample selection by AL and subsequent model adaptation by re-calibration dramatically reduces the overall error compared to the reference model (static ePLS). The improvements obtained through i) using ensembles rather than single models, ii) calibration set reweighing rather than operating on a sliding window and iii) explicit drift detection rather than continuous model adaptation are basically negligible considering the RMSE only (i.e. 2.1 vs. 2.08). However, from a practical point of view these extensions are able to massively reduce operational cost considering that AL might request a reference measurement at any time keeping operators basically locked at the process line when no implicit drift detection system is in place. On the other hand, in contrast to SW-PLS, the proposed method does not discard calibration samples, which over the long term reduces the bur-

Table 4: Cost-benefit tradeoff for different modelling and model adaptation variants. The proposed methods are highlighted in bold font.

Method/Criterion	RMSE local/global	Additional Parameters	# Reference Measurements
Static ePLS	2.75/3.29	None	None
SW-PLS [40]	-/2.49	Window Size	$\approx 12\%$
SW-PLS + AL [40]	-/2.1	Window Size	$\approx 12\%$
ePLS + AL	1.84/2.08	None	$8-10\%$ after drift
FLEXFIS-PLS (unsuperv.)	2.09/2.37	None	None

den of reference analyses especially in case that input-output relationships are subject to frequent, short-term alterations. Finally, unsupervised model adaptation using FLEXFIS-PLS could improve accuracy slightly when compared to static ePLS over all scenarios. However, unsupervised adaptation makes sense only if the drifted state is still represented well enough by the original latent variables, which is clearly not the case for scenario 1.

6. Conclusion

Calibration model maintenance remains a hot, yet in our opinion undervalued, issue in chemometrics and process analytical technology and currently involves large manual efforts and expert knowledge. Robust detection of the time-point when either the process itself or the process monitoring system fails is the first critical step towards autonomous, fully integrated maintenance of process reliability. Along these lines we here found that the Page-Hinkley statistic is a useful, general purpose quantity for tracking model diagnostic measures over time. In order to increase usability and flexibility of the PH test we introduced the concept of adaptive control limits, which circumvents using predefined thresholds (that need to be set according to application and quality measure). Committee disagreement significantly outperformed SoA indicators in terms of detection capability, delay of detection and false alarms indicating that the model diagnostic tool set should be extended by measures taking into account not only the "outlyingness" of test samples but

also the inference process. Once detected, drifts have to be compensated in a principled way with an emphasis on cost efficiency. Active learning in combination with automated model adaptation could meet these requirements while being remarkably robust against transient drift effects (outliers). Finally, the combination between fuzzy models and PLS is appealing since it allows both, fully unsupervised model adaptation and modelling non-linearity in the learning/mapping problem. Future work should focus on further integrating the here presented concepts. For instance, combining active learning with unsupervised model adaptation (i.e. semi-supervised adaptation) on the one hand and using the PH statistic to determine the time point when model adaptation can be safely shut down.

Acknowledgements

Financial support was provided by the Austrian research funding association (FFG) under the scope of the COMET program within the research project "Industrial Methods for Process Analytical Chemistry - From Measurement Technologies to Information Systems (imPACts)" (contract #843546) and by the strategic economic- and research program "Innovative Upper Austria 2020" of the province of Upper Austria. The third author was supported by the Basque Government through the BERC 2014-2017 program, and by the Spanish Ministry of Economy and Competitiveness MINECO through BCAM Severo Ochoa excellence accreditation SEV-2013-0323 and through project TIN2017-82626-R. This publication reflects only the authors' views.

References

[1] D. J. Merline, S. Vukusic, A. A. Abdala, Melamine formaldehyde: curing studies and reaction mechanism, Polym J 45 (4) (2013) 413–419.
URL http://dx.doi.org/10.1038/pj.2012.162

[2] H. P. Wohnsiedler, Polymerization in melamine-formaldehyde molded resins, Industrial & Engineering Chemistry 45 (10) (1953) 2307–2311.

arXiv:http://dx.doi.org/10.1021/ie50526a045, doi:10.1021/ie50526a045.
URL http://dx.doi.org/10.1021/ie50526a045

[3] M. Scheepers, R. Meier, L. Markwort, J. Gelan, D. Vanderzande, B. Kip, Determination of free melamine content in melamine-formaldehyde resins by raman spectroscopy, Vibrational Spectroscopy 9 (2) (1995) 139–146.

[4] M. Pawliczek, T. Reischer, J. Kowalski, W. Mrzinger, W. Kantner, M. Brandstetter, Long-term in-line monitoring of melamine formaldehyde resins using nir spectroscopy, Lenzinger Berichte 92 (2015) 47–52.

[5] C. Cernuda, E. Lughofer, P. Hintenaus, W. Märzinger, T. Reischer, M. Pawlicek, J. Kasberger, Hybrid adaptive calibration methods and ensemble strategy for prediction of cloud point in melamine resin production, Chemometrics and Intelligent Laboratory Systems 126 (2013) 60–75.

[6] B. Wise, R. Roginski, A calibration model maintenance roadmap, in: Proceedings of the 9th International Symposium on Advanced Control of Chemical Processes, IFAC, Whistler, British Columbia, Canada, 2015, pp. 260–265.

[7] M. Larrechi, M. Callao, Strategy for introducing nir spectroscopy and multivariate calibration techniques in industry, TrAC Trends in Analytical Chemistry 22 (9) (2003) 634 – 640. doi:http://dx.doi.org/10.1016/S0165-9936(03)01005-7.
URL http://www.sciencedirect.com/science/article/pii/S0165993603010057

[8] F. Serdio, E. Lughofer, A.-C. Zavoianu, K. Pichler, M. Pichler, T. Buchegger, H. Efendic, Improved fault detection employing hybrid memetic fuzzy modeling and adaptive filters, Applied Soft Computing 51 (2017) 60–82.

[9] I. Khamassi, M. Sayed-Mouchaweh, M. Hammami, K. Ghedira, Discussion and review on evolving data streams and concept drift adapting, Evolving Systems on-line and in press (2016) doi: 10.1007/s12530–016–9168–2.

[10] X. Wu, K.-H. Bellgardt, On-line fault detection of flow-injection analysis systems based on recursive next term parameter estimation, Analytica Chimica Acta 313 (3) (1995) 161–176.

[11] O. Haavisto, H. Hyotyniemi, Recursive multimodel partial least squares estimation of mineral flotation slurry contents using optical reflectance spectra, Analytica Chimica Acta 642 (2009) 102–109.

[12] C. Cernuda, E. Lughofer, G. Mayr, T. Röder, P. Hintenaus, W. Märzinger, J. Kasberger, Incremental and decremental active learning for optimized self-adaptive calibration in viscose production, Chemometrics and Intelligent Laboratory Systems 138 (2014) 14–29.

[13] F. Douak, F. Melgani, N. Alajlan, E. Pasolli, Y. Bazi, N. Benoudjit, Active learning for spectroscopic data regression, Journal of Chemometrics 26 (2012) 374–383.

[14] A. Ukil, J. Bernasconi, Neural network-based active learning in multivariate calibration, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42 (6) (2012) 1763–1771.

[15] C. Cernuda, E. Lughofer, H. Klein, C. Forster, M. Pawliczek, M. Brandstetter, Improved quantification of important beer quality parameters based on nonlinear calibration methods applied to ft-mir spectra, Analytical and Bioanalytical Chemistry 409 (3) (2017) 841–857.

[16] E. Lughofer, E. Weigl, W. Heidl, C. Eitzinger, T. Radauer, Recognizing input space and target concept drifts with scarcely labelled and unlabelled instances, Information Sciences 355–356 (2016) 127–151.

[17] A. Shaker, E. Hüllermeier, IBLStreams: a system for instance-based classification and regression on data streams, Evolving Systems 3 (2012) 239–249.

[18] P. Hintenaus, G. Kvas, W. Märzinger, An infrared spectrometer for process monitoring i, spectroscopy, in: Proceedings of the IEEE Industrial Electronics Society (IECON), Taipei, Taiwan, 2007.

[19] Z. Bleier, C. Brouillette, R. Carangelo, A monolithic interferometer for FT-IR spectroscopy, Spectroscopy 14 (10) (1999) 46–49.

[20] M. Haenlein, A. Kaplan, A beginner's guide to partial least squares (PLS) analysis, Understanding Statistics 3 (4) (2004) 283–297.

[21] B.-H. Mevik, V. Segtnan, T. Naes, Ensemble methods and partial least squares regression, Journal of Chemometrics 18 (11) (2004) 498–507.

[22] B. Efron, R. Tibshirani, An Introduction to the Bootstrap, Chapman and Hall/CRC, 1993.

[23] R. Rosipal, Kernel partial least squares for nonlinear regression and discrimination, Neural Network World 13 (3) (2003) 291–300.

[24] S. Wold, N. Kettaneh-Wold, B. Skagerberg, Nonlinear pls modeling, Chemometrics and Intelligent Laboratory Systems 7 (1) (1989) 53 – 65. doi:http://dx.doi.org/10.1016/0169-7439(89)80111-X.
URL http://www.sciencedirect.com/science/article/pii/016974398980111X

[25] S. Qin, T. McAvoy, Nonlinear pls modeling using neural networks, Computers & Chemical Engineering 16 (4) (1992) 379 – 391. doi:http://dx.doi.org/10.1016/0098-1354(92)80055-E.
URL http://www.sciencedirect.com/science/article/pii/009813549280055E

[26] W. Pedrycz, A. Skowron, V. Kreinovich, Handbook of Granular Computing, John Wiley & Sons, Chichester, West Sussex, England, 2008.

[27] T. Fearn, A. Davies, Locally-biased regression, Journal of Near Infrared Spectroscopy 11 (6) (2003) 467–478.

[28] E. Lughofer, C. Cernuda, S. Kindermann, M. Pratama, Generalized smart evolving fuzzy systems, Evolving Systems 6 (4) (2015) 269–292.

[29] E. Lughofer, FLEXFIS: A robust incremental learning approach for evolving TS fuzzy models, IEEE Transactions on Fuzzy Systems 16 (6) (2008) 1393–1410.

[30] P. F. Odgaard, B. Lin, S. B. Jorgensen, Observer and data-driven-model-based fault detection in power plant coal mills, IEEE Transactions on Energy Conversion 23 (2) (2008) 659–668. doi:10.1109/TEC.2007.914185.

[31] S. Qin, W. Li, H. Yue, Recursive PCA for adaptive process monitoring, Journal of Process Control 10 (5) (2000) 471–486.

[32] E. S. Page, Continuous inspection schemes, Biometrika 41 (1/2) (1954) 100–115. URL http://www.jstor.org/stable/2333009

[33] D. V. HINKLEY, Inference about the change-point from cumulative sum tests, Biometrika 58 (3) (1971) 509. doi:10.1093/biomet/58.3.509.

[34] A. Krogh, J. Vedelsby, Neural network ensembles, cross validation, and active learning, in: Advances in Neural Information Processing Systems, MIT Press, 1995, pp. 231–238.

[35] J. Gama, Knowledge Discovery from Data Streams, Chapman & Hall/CRC, Boca Raton, Florida, 2010.

[36] H. Wang, Z. Abraham, Concept drift detection for streaming data, in: Proc. of the International Joint Conference of Neural Networks (IJCNN) 2015, Killarney, Ireland, 2015, pp. 1–9.

[37] A. Shaker, E. Lughofer, Self-adaptive and local strategies for a smooth treatment of drifts in data streams, Evolving Systems 5 (4) (2014) 239–257.

[38] N. K. M. Faber, R. Bro, Standard error of prediction for multiway pls: 1. background and a simulation study, Chemometrics and Intelligent Laboratory Systems 61 (12) (2002) 133 – 149. doi:https://doi.org/10.1016/S0169-7439(01)00204-0. URL http://www.sciencedirect.com/science/article/pii/S0169743901002040

[39] W. Ni, S. D. Brown, R. Man, Stacked partial least squares regression analysis for spectral calibration and prediction, Journal of Chemometrics 23 (10) (2009) 505–517.

624 doi:10.1002/cem.1246.
625 URL http://dx.doi.org/10.1002/cem.1246

626 [40] C. Cernuda, E. Lughofer, T. Reischer, W. Kantner, M. Pawliczek, M. Brandstaetter, Dynam-
627 ically slided chemometric models for robust on-line prediction of cloud point in melamine
628 resin production, in: Proceedings of the Conferentia Chemometrica 2015, Budapest, Hun-
629 gary, 2015, downloadable at http://cc2015.ttk.mta.hu/.